amv_mfo_control_g9.tex

Yet again on iteration improvement for averaged expected cost control for 1D ergodic diffusions

S.V. Anulova*[†], H. Mai[‡§], A.Yu. Veretennikov[¶‖]

September 2, 2017

Abstract

An ergodic Hamilton-Jacobi-Bellman (HJB) equation is proved for a 1D controlled diffusion with variable diffusion and drift coefficients both depending on control. Also, convergence of the iteration improvement algorithm to its (unique) solution is established.

*Institute for Control Sciences, Moscow, Russia; email: anulovas @ ipu.ru

†For the first author this research has been supported by the Russian Foundation for Basic Research grant no. 16-08-01285

‡CREST and ENSAE ParisTech, France; email: hilmar.mai @ gmail.com

§The second author thanks the Institut Louis Bachelier for financial support.

¶University of Leeds, UK, & National Research University Higher School of Economics, & Institute for Information Transmission Problems, Moscow, Russia; email: a.veretennikov @ leeds.ac.uk

‖The third author is grateful to the financial support by the DFG through the CRC 1283 "Taming uncertainty and profiting from randomness and low regularity in analysis, stochastics and their applications" at Bielefeld University and for highly stimulating atmosphere during his stay there in August 2017; also, for this author this study has been funded by the Russian Academic Excellence Project '5-100' and by the Russian Foundation for Basic Research grant no. 17-01-00633_a. All the authors gratefully acknowledge the support and hospitality of the Oberwolfach Research Institute for Mathematics (MFO) during the RiP programme in June 2014 where this study was initiated.

1 Introduction

The area of controlled diffusion processes include various settings: probably, the most developed is the directions of the finite horizon one (in turn, split into some futher sub-areas which we do not discuss); infinite horizon settings are "split" into discount long-term control, averaged long-term control and total expected long-term control. The present paper relates to the averaged long-term control.

Discrete time controlled models with averaged ergodic control were considered in the monographs [5], [7, 8], [14], [18] and in some others; many important journal references can be found therein.

For continuous time, Mandl establishedin his book [13] apparently first results on ergodic averaged control for controlled 1D diffusion on a finite interval with boundary conditions including jumps from the boundary. He established the HJB (Bellman's) equation and proved uniqueness of the couple up to a constant for the first component. Improvement of control was also briefly discussed, however, without convergence.

Morton [15] considered a 1D case (in fact, a multi-dimensional case was treated in this paper, too, under somewhat stronger assertions: we do not touch multi-dimensional processes in the present paper), with a price function defined by (5) without any relation to (3) (see below for both). Morton proved ([15, Theorem 1]) that the optimal price does satisfy an ergodic Bellman's equation and also that the policy determined by argsup in Bellman's equation is optimal within some rather special class of Markov policies which are fixed functions outside some bounded interval; he also established a certain inequality for the optimal price and for any solution of Bellman's equation; neither uniqueness for the solution of Bellman's equation was established, nor convergence of the RIA (reward improvement algorithm, see below) towards a solution, although, the RIA itself has been mentioned.

Later in the 80s the monograph [3] considers ergodic control for diffusion processes. In the works [1] and [2], 1D diffusions are treated among other settings with "relaxed control", under certain recurrence assumptions and under two types of condition, "stable" or "near-monotone". For some technical reasons, control in the diffusion coefficient is not allowed. In this setting, the paper establishes Bellman's equation, existence, uniqueness, and convergence of the RIA.

In the present paper, a 1D diffusion is considered for which both drift and diffusion coefficients may depend on the control; the latter is understood as a "Markov feedback" one. So called relaxed control is not considered here. In this paper we only study weak solutions of SDEs. Although, the ideas for both – strong and weak solutions – cases are similar, it is difficult to combine them both in one text. Recall that weak SDE solution assume that, generally speaking, for each new SDE with a new pair of coefficients, it may be required to change Wiener process and possibly the whole probability space as well. However, we never compare the trajectories of two solutions in one formula, so that there is no confusion about a probability space. In fact, the authors' view is that control for strong solutions is, of course, more sensible because there is no need to think what does it mean that probability spaces may change. Yet, controlled weak solutions are also popular in the literature, and it turns out that some technicalities are easier in this case. This is why the weak solution setting has been chosen for this paper. Another article will be devoted to strong solutions in the near future.

The whole idea of this paper for either setting, weak or strong, relates also to the presentation in [10, Ch.1, section 1] where one of the explanations of the idea of Bellman's equation is given deliberately a bit non-rigorously, on some example related to the infinite horizon control. This paper is an attempt to make this presentation rigorous by changing the setting, at the price of losing some of the simplicity of course. However, it is worth mentioning that in the "more standard" finite horizon or infinite horizon with discount settings, the class of admissible strategies is wider, although it is known that feedback or "Markov" strategies are sufficient. In the ergodic averaged control, we only consider *homogeneous* feedback strategies and do not allow any wider class of strategies.

The paper consists of five sections: 1 – Introduction, 2 – Setting, 3 – Assumptions and Auxiliaries, 4 – Main result(s) and 5 – Proof(s).

2 Setting

Given a standard probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ and a one-dimensional $(\mathcal{F}_t)$ Wiener process $B = (B_t)_{t\geq 0}$ on it we consider a one-dimensional SDE with coefficients b, σ and a control parameter α described as follows:

$$\begin{aligned} dX_t^\alpha &= b(\alpha(X_t^\alpha), X_t^\alpha)\, dt + \sigma(\alpha(X_t^\alpha), X_t^\alpha)\, dW_t, \quad t \geq 0, \\ X_0^\alpha &= x \in \mathbb{R}. \end{aligned} \tag{1}$$

Its (weak) solution does exist [11] and under our conditions – 1D, boundedness of all coefficients and uniform non-degeneracy (or ellipticity) of σ^2 – is weakly unique.

Let a non-empty compact set $U \subset \mathbb{R}$ be a range of possible control values. Without any further reminder U being compact is always bounded. Let $b : U \times \mathbb{R} \to \mathbb{R}$,

$\sigma : U \times \mathbb{R} \to \mathbb{R}$, $\alpha : \mathbb{R} \to U$ be given Borel functions (some more regularity assumptions will be presented later).

Denote the (extended) generator, which corresponds to the equation (1) with a fixed function $\alpha(\cdot)$ by L^α :

$$L^\alpha(x) = b(\alpha(x), x)\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2(\alpha(x), x)\frac{\partial^2}{\partial x^2}, \quad x \in \mathbb{R}.$$

Given a running cost function $f : U \times \mathbb{R} \to \mathbb{R}$ from a suitable function class we aim to choose an optimal (in some relaxed setting, at least, "nearly-optimal") control strategy $\alpha : \mathbb{R} \to U$ (Markov homogeneous, or, in another language, Markov feedback strategy) such that the corresponding solution X^α minimizes the averaged cost function

$$\rho^\alpha(x) := \limsup_{T\to\infty} \frac{1}{T}\int_0^T \mathbb{E}_x f(\alpha(X_t^\alpha), X_t^\alpha)\, dt. \tag{2}$$

By $\mathcal{K}$ we denote the class of strategies $\alpha : \mathbb{R} \to U$ which are Borel measurable. For convenience for every $\alpha \in \mathcal{K}$ we define the function $f^\alpha : \mathbb{R} \to \mathbb{R}$, $f^\alpha(x) = f(\alpha(x), x)$, $x \in \mathbb{R}$. Now, instead of(2) we can use the equivalent form,

$$\rho^\alpha(x) = \limsup_{T\to\infty} \frac{1}{T}\int_0^T \mathbb{E}_x f^\alpha(X_t^\alpha)\, dt.$$

Finally, the "minimax" cost function is defined by the expression

$$\rho(x) := \inf_{\alpha\in\mathcal{K}} \limsup_{T\to\infty} \frac{1}{T}\int_0^T \mathbb{E}_x f^\alpha(X_t^\alpha)\, dt. \tag{3}$$

Suppose that for every $\alpha \in \mathcal{K}$ the solution of the equation (1) X^α is an ergodic process, that is, there exists a unique limiting distribution μ^α of X_t^α, $t \to \infty$, the same for all initial conditions $X_0 = x \in \mathbb{R}$. Then it is true that for every $x \in \mathbb{R}$,

$$\rho^\alpha(x) \equiv \rho^\alpha := \int f^\alpha(x)\, \mu^\alpha(dx) =: \langle f^\alpha, \mu^\alpha \rangle, \tag{4}$$

and

$$\rho(x) \equiv \rho := \inf_{\alpha\in\mathcal{K}} \int f^\alpha(x)\, \mu^\alpha(dx) = \inf_{\alpha\in\mathcal{K}} \langle f^\alpha, \mu^\alpha \rangle. \tag{5}$$

(Note that under our assumptions ρ does not depend on x.) Ergodicity requires special conditions on the characteristics b, σ, α; they will be later specified in the next section. (Here we do not want to restrict ourselves to particular technical assumptions because it is likely that they may allow certain variability.) We also define

$$v^\alpha(x) := \int_0^\infty E_x(f^\alpha(X_t^\alpha) - \rho^\alpha)\, dt, \quad \alpha \in \mathcal{K}.$$

This integral will converge under the recurrence assumptions below.

Solutions of the equation (1) will be understood as weak ones. Correspondlingly, the ergodic HJB equation (6) – see the next paragraph – will be established for the weak solution setting, altough it actually holds for both; however, strong SDE solutions will be treated separately in anoter paper. The main difference is that in the strong setting, the optimal strategy may not exist, which could be considered as a bit less convenient; one more difference lies in the convergence issues. On the other hand, strong SDE solutions are, of course, more natural.

The first goal of the paper is to prove that the cost ρ – which is a constant in the ergodic setting – is the component of the pair (V, ρ), which is a unique solution of the *ergodic HJB or Bellman's equation,*

$$\inf_{u \in U} [L^u V(x) + f^u(x) - \rho] = 0, \quad x \in \mathbb{R}, \tag{6}$$

where V will be unique up to an additive constant, while ρ will be unique in the standard sense. The meaning of the function V is that it coincides with v^α for the optimal strategy α if the latter exists, and this function is the main tool for finding the optimal or close to optimal strategies.

The second goal is to show that the "RIA" algorithm ("reward improvement algorithm", or, in some papers, "PIA" for "policy improvement algorithm") provides a sequence of convergent approximate costs, $\rho_n \to \rho,\ n \to \infty$. It is interesting that under our minimal assumptions of regularity on strategies for the weak setting it is yet possible to justify a monotonic convergence $\rho_n \downarrow \rho,\ n \to \infty$, of the "exact" RIA, unlike for the strong setting; compare to [10, Krylov, ch.1, §4] it looks necessary – or, at least, natural – to work with "approximate" RIA (called Bellman–Howard's iteration procedure there) and regular – Lipschitz – strategies.

Concerning equation (6), it may look like it lacks some boundary conditions: indeed, a 2nd order PDE normally does require some boundary conditions, which, for example, in the considered 1D case simply mean two boundary conditions at two end-points if the equation is on a bounded interval. However, this is the equation "on the whole space" and we are going to solve it in a specific class of functions V – namely, bounded, or, at most, moderately growing, – which in some sense substitutes the (Dirichlet) boundary conditions at $\pm\infty$. In fact, this situation is similar to the non-controlled setting of the Poisson equation "on the whole space", see, e.g., [16].

Also let us emphasize that unlike in the finite horizon case, here in the average ergodic control setting, the solution of the HJB equation is *a couple* (V, ρ), where ρ is the desired cost while V is some auxiliary function, which also admits a certain interpretation in terms of control theory.

Concerning uniqueness, we note that, clearly, with any couple (V, ρ) and any constant C, the couple $(V + C, \rho)$ is also a solution. There are two close enough to each other options how to tackle this fact: either accept that uniqueness will be established up to a constant, or to choose a certain "natural" constant satisfying some "centering condition" as will be done below. In any case, for computing the optimal strategy this additive constant is of no importance because in any case and for any control $L^u V = L^u(V + C)$.

3 Assumptions and auxiliaries

To ensure ergodicity of X^α under any feedback control strategy $\alpha \in K$, we make the following assumptions on the drift and diffusion coefficients.

(A1) The function b is bounded, C^1 in x, and

$$\lim_{|x|\to\infty} \sup_{u\in U} x\, b(u, x) = -\infty. \tag{7}$$

(A2) The function σ is bounded, uniformly non-degenerate and C^1 in x.

(A3) The function f is bounded.

(A4) The function σ is continuous in u.

(A5) The set U is compact.

The assumption (A1) may be relaxed a little bit. The assumption (A5) is required only so as to attain the inf values in the expressions like (6).

We will need the following

Lemma 1. *Let the assumptions (A1) – (A4) be satisfied. Then the cost function v^α has the following properties:*

1. *There exist* $C, m > 0$ *such that* $\sup_\alpha |v^\alpha|(x) \le C(1+|x|^m)$.

2. v^α *is continuous.*

3. $v^\alpha \in W^2_{p,loc}$ *for any* $p \ge 1$.

4. v^α *satisfies a Poisson equation in the whole space,*

$$L^\alpha v^\alpha(x) + f^\alpha(x) - < f^\alpha, \mu^\alpha >= 0. \qquad (8)$$

5. *Solution of this equation is unique up to an additive constant in the class of Sobolev solution with a no more than some (any) polynomial growth for* v^α.

6. $< v^\alpha, \mu^\alpha >= 0$.

Note that the condition 3 in the 1D case is equivalent to saying that $v^\alpha, (v^\alpha)', (v^\alpha)'' \in L_{p,loc}$ for any $p \ge 1$). In higher dimensions there is no such simple equivalence.

Proof. All claims follow from [20] and [17]; see also [9, Lemma 4.13 and Remark 4.3]. □

Lemma 2. *Let the assumptions (A1) – (A3) hold true. Then,*

- *For any* $C_1, m_1 > 0$ *there exist* $C, m > 0$ *such that for any strategy* $\alpha \in \mathcal{K}$ *and for any function* g *growing no faster than* $C_1(1+|x|^{m_1})$,

$$\sup_t |\mathbb{E}_x g(X^\alpha_t)| \le C(1+|x|^m). \qquad (9)$$

- *For any strategy* $\alpha \in \mathcal{K}$ *the function* ρ^α *is a constant, and there exists* $C < \infty$ *such that*

$$\sup_\alpha |\rho^\alpha| \le C < \infty. \qquad (10)$$

- *For any* $\alpha \in \mathcal{K}$, *the invariant measure* μ^α *integrates any polynomial:*

$$\int |x|^m \, \mu^\alpha(dx) < \infty.$$

Proof follows from [20] and [17].

Remark 1. *Note that because of $D = 1$, under the assumptions (A1)–(A2) for any $\alpha \in \mathcal{K}$ there is a unique stationary measure μ^α, and it is* equivalent to the Lebesgue measure Λ. *The latter follows from the well-known explicit representation of the stationary density (here $\bar b(x) = b(\alpha(x), x)$ for some – any – fixed α),*

$$f(x) = \frac{c}{\sigma^2(x)}\ \exp(\int_0^x (2\bar b/\sigma^2)(y)\,dy),$$

with a normalizing constant c. Recall that it may be easily checked, e.g., as follows. Let $v(x) = \dfrac{c}{2}\exp(\displaystyle\int_0^x (2\bar b/\sigma^2)(y)\,dy)$. Since $\sigma^2 f(x)/2 = v(x)$, and $v'(x) = \frac{2\bar b}{\sigma^2}v(x) = \bar b f(x)$, we find,

$$L^* f(x) = \frac{1}{2}(\sigma^2(x)f(x))'' - (\bar b(x)f(x))' = v(x)'' - (v'(x))' = 0,$$

as required.

4 Main result

Recall that the state space dimension is $D = 1$ and that all SDE solutions with any Markov strategy may be weak (although, if by any chance some of them are strong, it would not be a contradiction); we want them to be unique in distribution, strong Markov and ergodic. All of these follow from [11] and from the assumptions (A1) and (A2) (see [20] about ergodicity).

The “exact RIA” reads as follows. Let us start with some homogeneous Markov strategy α_0, which uniquely determines $\rho_0 = \rho^{\alpha_0} \equiv \langle f^{\alpha_0}, \mu^{\alpha_0}\rangle$ and $v_0 = v^{\alpha_0}$. Next, for any couple (v, ρ) such that $v \in C^2$, or $v \in W^2_{p,loc}$ with any $p > 0$, and for $\rho \in \mathbb{R}$, define

$$F[v,\rho](x) := \inf_{u\in U}\left[L^u v(x) + f^u(x) - \rho\right].$$

Now, by induction given α_n, ρ_n and v_n, the next “improved” strategy α_{n+1} is defined as follows: for any x,

$$L^{\alpha_{n+1}} v_n(x) + f^{\alpha_{n+1}}(x) - \rho_n = F[v_n, \rho_n](x). \qquad (11)$$

which is equivalent to

$$L^{\alpha_{n+1}} v_n(x) + f^{\alpha_{n+1}}(x) = \inf_u\left[L^u v_n(x) + f^u(x)\right] =: G[v_n](x).$$

We assume that a Borel measurable version of such a strategy may be chosen. In our situation for each x the set $\mathrm{Arginf}_u(F^u[v,\rho](x))$ is closed. In this case the existence of such Borel strategy is proved in [6, Remark 3, Proposition 2, & Theorem 5] with a reference to earlier papers, or in [4, Theorem 1] where it is called Stschegolkow's Theorem with a reference to [12, Theorem 39]. Note that by some reason the name of the second author of [12] in the reference [4] is missing; in turn, in [6] the reference to the paper [19] is given with a wrong year (1943 instead of the correct 1948). In any case, the result in question is, apparently, due to E.A. Schegolkov (Stschegolkow) [19].

The value ρ_{n+1} is then defined as

$$\rho_{n+1} := \langle f^{\alpha_{n+1}}, \mu^{\alpha_{n+1}} \rangle,$$

where, in turn, $\mu^{\alpha_{n+1}}$ is the (unique) invariant measure, which corresponds to the strategy α_{n+1}. Recall that

$$v_n(x) = \int_0^\infty \mathbb{E}_x(f^{\alpha_n}(X_t^{\alpha_n}) - \rho_n)\, dt,$$

and

$$v_{n+1}(x) = \int_0^\infty \mathbb{E}_x(f^{\alpha_{n+1}}(X_t^{\alpha_{n+1}}) - \rho_{n+1})\, dt.$$

Theorem 1. *Let the assumptions (A1) – (A5) be satisfied. Then:*

1. For any n, $\rho_{n+1} \le \rho_n$, the sequence ρ_n is bounded, and there is a limit $\rho_n \downarrow \tilde\rho$.

2. The sequence (v_n) is tight (pre-compact) in $C^2[\ell_1, \ell_2]$ on any interval $[\ell_1, \ell_2]$, and there exists a limit $\lim_{n\to\infty} v_n(x) =: \tilde v(x)$.

3. The couple $(\tilde v, \tilde\rho)$ solves the equation (6); also, $\tilde v'' \in Lip$ locally.

4. This solution is unique in the class of functions growing no faster than some (any) polynomial and belonging to the class $W^2_{p,loc}$ for any $p > 0$.

5. The component $\tilde\rho$ in the couple $(\tilde v, \tilde\rho)$ coincides with ρ.

6. If there is a couple $(\hat V(x), \hat\rho)$ with $\hat V$ growing no faster than some polynomial, such that $\sup_x |F[\hat V, \hat\rho](x)| \le \epsilon$, then $|\rho - \hat\rho| \le \epsilon$.

5 Proof

1. Due to (11) and (8),

$$\rho_n = \rho^{\alpha_n} = \langle f^{\alpha_n}, \mu^{\alpha_n} \rangle \stackrel{a.s.}{=} L^{\alpha_n} v_n(x) + f^{\alpha_n}(x)$$

$$\ge \inf_u [L^u v_n(x) + f^u(x)] = L^{\alpha_{n+1}} v_n(x) + f^{\alpha_{n+1}}(x),$$

and

$$\rho_{n+1} \stackrel{a.s.}{=} L^{\alpha_{n+1}} v_{n+1}(x) + f^{\alpha_{n+1}}(x).$$

So,

$$\rho_n - \rho_{n+1} \stackrel{a.s.}{\geq} (L^{\alpha_{n+1}} v_n + f^{\alpha_{n+1}})(x) - (L^{\alpha_{n+1}} v_{n+1} + f^{\alpha_{n+1}})(x)$$

$$= (L^{\alpha_{n+1}} v_n - L^{\alpha_{n+1}} v_{n+1})(x).$$

Now let us apply Ito's formula with expectations (also known as Dynkin's formula) to $(v_n - v_{n+1})(X_t^{\alpha_{n+1}})$:

$$\mathbb{E}_x v_n(X_t^{\alpha_{n+1}}) - \mathbb{E}_x v_{n+1}(X_t^{\alpha_{n+1}}) - v_n(x) + v_{n+1}(x)$$

$$= \mathbb{E}_x \int_0^t (L^{\alpha_{n+1}} v_n - L^{\alpha_{n+1}} v_{n+1})(X_s^{\alpha_{n+1}})\, ds$$

$$\leq \mathbb{E}_x \int_0^t (\rho_n - \rho_{n+1})\, ds = (\rho_n - \rho_{n+1})\, t.$$

Since the first term $\mathbb{E}_x v_n(X_t^{\alpha_{n+1}}) - \mathbb{E}_x v_{n+1}(X_t^{\alpha_{n+1}}) - v_n(x) + v_{n+1}(x)$ in this double (in)equality is bounded for a fixed x, divide all terms of the latter inequality by t and let $t \to \infty$. We get,

$$0 \leq \rho_n - \rho_{n+1},$$

as required.

Hence, $\rho_n \geq \rho_{n+1}$, so that $\rho_n \downarrow \tilde{\rho}$ (since ρ_n is bounded from below for bounded f^α) with some $\tilde{\rho}$. Thus, **the RIA does converge**, although so far we do not know whether $\tilde{\rho} = \rho$.

Note that clearly $\tilde{\rho} \geq \rho$, since ρ is the inf over all Markov strategies, while $\tilde{\rho}$ is the inf over some countable subset (a sequence) of them. *We shall see later that they coincide: $\tilde{\rho} = \rho$.*

Recall that now we want to show that $v_n \to \tilde{v}$ such that the couple $(\tilde{v}, \tilde{\rho})$ satisfies the HJB equation (6), and that $\tilde{\rho}$ – as well as $\tilde{v}$ in some sense – here is unique.

2. What we want to do is to pass to the limit in the equation

$$L^{\alpha_{n+1}} v_{n+1}(x) + f^{\alpha_{n+1}}(x) - \rho_{n+1} = 0, \quad \text{as } n \to \infty,$$

after having showed compactness of the set (v_n) at least in C^1 (and later on in C^2). Denote

$$G[v_n] := \inf_{\alpha}(L^\alpha v_n + f^\alpha) \quad (\equiv L^{\alpha_{n+1}} v_n + f^{\alpha_{n+1}}),$$

and

$$F_1[x, v', \rho] := \inf_u [\hat{b}^u v' + \hat{f}^u - \hat{\rho}](x) \equiv \inf_u \left[\hat{b}^u v' + \hat{f}^u - \frac{\rho}{a^u}\right](x),$$

where

$$a^u(x) = \frac{1}{2}(\sigma^u(x))^2, \quad \hat{b}^u(x) = b^u(x)/a^u(x),$$

$$\hat{f}^u(x) = f^u(x)/a^u(x), \quad \hat{\rho}^u(x) = \rho/a^u(x).$$

Note that due to non-degeneracy of σ^2, the equation (slightly abusing notations, we stop repeating that in most cases equalities are a.s.)

$$L^{\alpha_{n+1}} v_{n+1}(x) + f^{\alpha_{n+1}}(x) - \rho_{n+1} = 0 \tag{12}$$

implies uniform boundedness of (v_n'') (so, in particular, (v_n) is tight, at least, in C^1 on any bounded interval in $\mathbb{R}$). Also, we have,

$$G[v_n] - \rho_n = L^{\alpha_{n+1}} v_n(x) + f^{\alpha_{n+1}}(x) - \rho_n.$$

Further, note that

$$G[v_n] - \rho_n = \inf_{\alpha}[L^\alpha v_n(x) + f^\alpha(x) - \rho_n] \le 0.$$

By subtracting zero (12) from $G[v_n] - \rho_n$, we obtain

$$G[v_n] - \rho_n = L^{\alpha_{n+1}}(v_n(x) - v_{n+1}(x)) - (\rho_n - \rho_{n+1}). \tag{13}$$

3. Now we **want to show** that

$$\tilde{v}'(x) - \tilde{v}'(r) + \int_r^x F_1[s, \tilde{v}'(s), \tilde{\rho}]\, ds = 0, \tag{14}$$

which implies by differentiation that

$$\tilde{v}''(x) + F_1[x, \tilde{v}', \tilde{\rho}](x) = 0. \tag{15}$$

Note that, in fact, this equation is equivalent to the HJB equation (6).

Let us now show why (13) indeed, implies (14) and, hence, also (15). Let us divide all terms in (13) by $a_{n+1} := a^{\alpha_{n+1}}$: we obtain,

$$0 \ge \frac{(G[v_n](x) - \rho_n)}{a_{n+1}}$$

$$= (v_n''(x) - v_{n+1}''(x)) + (\hat{b}^{\alpha_{n+1}}(v_n' - v_{n+1}')) - \frac{(\rho_n - \rho_{n+1})}{a_{n+1}}$$

$$\ge (v_n''(x) - v_{n+1}''(x)) - \frac{K}{\delta}|v_n'(x) - v_{n+1}'(x)| - \frac{1}{\delta}(\rho_n - \rho_{n+1}). \tag{16}$$

So, we have already established that

$$(v_n''(x) - v_{n+1}''(x)) - \frac{K}{\delta}|v_n'(x) - v_{n+1}'(x)| - \frac{\rho_n - \rho_{n+1}}{\delta} \le 0. \tag{17}$$

The next trick is to note that again due to (16)

$$0 \ge G[v_n] - \rho_n \ge a_{n+1}(v_n'' - v_{n+1}'') - C'|v_n' - v_{n+1}'| - (\rho_n - \rho_{n+1}),$$

which after division by a_{n+1} implies that

$$0 \ge v_n'' + F_1[v_n', \rho_n] \ge ((v_n'' - v_{n+1}'') - C|v_n' - v_{n+1}'|) - C(\rho_n - \rho_{n+1}). \tag{18}$$

Now integrating (18) with a note that (v_n) **is tight in** C^1, – whence, $v_{n'}' \to \tilde{v}'$ and $v_{n'+1}' \to \tilde{v}'$ in C on any finite interval over some subsequence $n' \to \infty$, – we get with $x > r$,

$$0 \ge v_{n'}'(x) - v_{n'}'(r) + \int_r^x F_1[s, v_{n'}'(s), \rho_{n'}]\, ds - C(\rho_{n'} - \rho_{n'+1})(x - r)$$

$$\ge \int_r^x ((v_{n'}'' - v_{n'+1}'')(s) - C|v_{n'}' - v_{n'+1}'|(s))\, ds - C(\rho_{n'} - \rho_{n'+1})(x - r)$$

$$= v_{n'}'(x) - v_{n'}'(r) - v_{n'+1}'(x) + v_{n'+1}'(r)$$

$$-C \int_r^x |v_{n'}' - v_{n'+1}'|(s)\, ds - C(\rho_{n'} - \rho_{n'+1})(x - r).$$

Hence, we have established the following double inequality,

$$0 \ge v_n'(x) - v_n'(r) + \int_r^x F_1[v_n'(s), \rho_n] - C(\rho_{n'} - \rho_{n'+1})(x - r)$$

$$\geq v_n'(x) - v_n'(r) - v_{n+1}'(x) + v_{n+1}'(r)$$
$$-C\int_r^x |v_n' - v_{n+1}'|(s)ds - C(\rho_{n'} - \rho_{n'+1})(x-r). \tag{19}$$

Since $v_{n'} \to \tilde{v}$ and $v_{n'}' \to \tilde{v}'$ in C on any bounded interval, and since $\rho_n \downarrow \tilde{\rho}$ and, hence, $\rho_n - \rho_{n+1} \to 0$, we get from the inequality (19) that in the limit as $n' \to \infty$:

$$0 \geq \tilde{v}'(x) - \tilde{v}'(r) + \limsup_{n'\to\infty} \int_r^x F_1[v_{n'}'(s), \rho_{n'}]\, ds$$
$$\geq \tilde{v}'(x) - \tilde{v}'(r) + \liminf_{n'\to\infty} \int_r^x F_1[v_{n'}'(s), \rho_{n'}]\, ds \geq 0,$$

the latter inequality due to the fact that the right hand side in (19), clearly, goes to zero. Note for the sequel that the limit $\tilde{v}$ is necessarily bounded along with its derivative $\tilde{v}'$ on any bounded interval.

Here $F_1[s, v_n'(s), \rho_n] \to F_1[s, v'(s), \tilde{\rho}]$, because $v_n'(s) \to v'(s)$ and $\rho_n \to \tilde{\rho}$, while $F_1[\xi, \eta]$ is continuous (and even Lipschitz continuous) in (ξ, η), since

$$|F_1[s, \xi, \eta] - F_1[s, \xi', \eta']|$$
$$= |\inf_u \left[\frac{b^u(s)}{a^u(s)}\xi + \frac{f^u(s)}{a^u(s)} - \frac{\eta}{a^u(s)}\right] - \inf_u \left[\frac{b^u(s)}{a^u(s)}\xi' + \frac{f^u(s)}{a^u(s)} - \frac{\eta'}{a^u(s)}\right]|$$
$$\leq \sup_u \frac{b^u(s)}{a^u(s)}|\xi - \xi'| + \frac{|\eta - \eta'|}{a^u(s)} \leq C(|\xi - \xi'| + |\eta - \eta'|).$$

So, from (19) we obtain the desired equation (14). This, in turn, implies $\tilde{v} \in C_b^2$, and by differentiation we get the equation (15) **equivalent to the HJB equation (6)**,

$$\tilde{v}''(x) + F_1[x, \tilde{v}', \tilde{\rho}](x) = 0. \tag{20}$$

Also, it follows that $\tilde{v}''$ is bounded on any bounded interval.

4. Further, $F_1[s, \tilde{v}', \tilde{\rho}](s)$ is locally Lipschitz in s. Indeed, since $\tilde{v}'$ is Lipschitz and since both $\tilde{v}'$ and $\tilde{v}''$ are bounded on any bounded interval, for any $N > 0$ we have for $|s|, |s'| \le N$,

$$|F_1[s, \tilde v'(s), \tilde\rho(s)] - F_1[s', \tilde v'(s'), \tilde\rho(s')]|$$

$$= |\inf_u \left[\frac{b^u(s)}{a^u(s)}\tilde v'(s) + \frac{f^u(s)}{a^u(s)} - \frac{\tilde\rho}{a^u(s)}\right] - \inf_u \left[\frac{b^u(s')}{a^u(s')}\tilde v'(s') + \frac{f^u(s')}{a^u(s')} - \frac{\tilde\rho}{a^u(s')}\right]|$$

$$\le \sup_u(|\frac{b^u(s)}{a^u(s)}\tilde v'(s) - \frac{b^u(s')}{a^u(s')}\tilde v'(s')| + |\frac{\tilde\rho}{a^u(s)} - \frac{\tilde\rho}{a^u(s')}|) \le C_N|s - s'|.$$

Hence, it follows from (20) that also locally,

$$\tilde v'' \in \mathrm{Lip},$$

as required.

5. *Uniqueness for ρ.* Suppose there are two solutions of the (HJB) equation, v^1, ρ^1 and v^2, ρ^2 with a polynomial growth for v^i. Denote $v(x) := v^1(x) - v^2(x)$ and consider two strategies $\alpha_1(x) = \mathrm{argsup}_u(L^u v(x))$ and $\alpha_2(x) = \mathrm{arginf}_u(L^u v(x))$, and denote by X^i_t a (weak) solution of the SDE corresponding to each strategy α_i. (It exists and is weakly unique.)

Note that

$$h_2(x) := \inf_u(L^u v(x) + \rho^1 - \rho^2)$$

$$= \inf_u(L^u v^1(x) + f^u(x) + \rho^1 - (L^u v^2(x) + f^u(x) + \rho^2))$$

$$\le \inf_u(L^u v^1(x) + f^u(x) + \rho^1) - \inf_u(L^u v^2(x) + f^u(x) + \rho^2) = 0,$$

and similarly,

$$h_1(x) := \sup_u(L^u v(x) + \rho^1 - \rho^2)$$

$$= -\inf_u(L^u(-v)(x) + \rho^2 - \rho^1)$$

$$= -\inf_u(L^u v^2(x) + f^u(x) + \rho^2 - (L^u v^1(x) + f^u(x) + \rho^1))$$

$$\ge -[\inf_u(L^u v^2(x) + f^u(x) + \rho^2) - \inf_u(L^u v^1(x) + f^u(x) + \rho^1)] = 0.$$

We have,

$$L^{\alpha_2} v(x) = h_2(x) + \rho^2 - \rho^1,$$

and

$$L^{\alpha_1} v(x) = h_1(x) + \rho^2 - \rho^1.$$

Further, Ito's (Dynkin's) formula is applicable. So,

$$\mathbb{E}_x v(X_t^1) - v(x) = \mathbb{E}_x \int_0^t L^{\alpha_1} v(X_s^1)\, ds$$

$$= \mathbb{E}_x \int_0^t h_1(X_s^1)\, ds + (\rho^2 - \rho^1)\, t \stackrel{(h_1 \geq 0)}{\geq} (\rho^2 - \rho^1)\, t.$$

Here the left hand side is bounded (x fixed) due to the Lemma 2, so, we obtain,

$$\rho_1 - \rho_2 \geq 0.$$

Similarly, considering α_2 and using $h_2 \leq 0$ we conclude that

$$\mathbb{E}_x v(X_t^2) - v(x) = \mathbb{E}_x \int_0^t L^{\alpha_2} v(X_s^2)\, ds$$

$$= \mathbb{E}_x \int_0^t h_2(X_s^2)\, ds + (\rho^2 - \rho^1)\, t;$$

so, due to the boundedness of the left hand side (Lemma 2),

$$\rho^1 - \rho^2 = \liminf_{t \to 0} (t^{-1} \mathbb{E}_x \int_0^t h_2(X_s^2)\, ds) \stackrel{(h_2 \leq 0)}{\leq} 0.$$

Thus, eventually,

$$\rho_1 = \rho_2.$$

6. *Proof of the equality* $\rho = \tilde{\rho}$. We have seen that for any initial (α_0, ρ_0), the sequence ρ_n converges monotonically to $\tilde{\rho}$, which is a unique component of solution of the Bellman equation (6).

Hence, given some (any) $\epsilon > 0$, take any initial strategy α_0 such that

$$\rho_0 = \rho^{\alpha_0} < \rho + \epsilon.$$

Then, clearly, the corresponding limit $\tilde{\rho}$ will satisfy the same inequality,

$$\tilde{\rho} = \lim_{n} \rho_n < \rho + \epsilon.$$

Due to uniqueness of $\tilde{\rho}$ as a component of solution of the equation (6) and since $\epsilon > 0$ is arbitrary, we conclude that

$$\tilde{\rho} \le \rho.$$

But also $\tilde{\rho} \ge \rho$. So,

$$\tilde{\rho} = \rho.$$

7. *Uniqueness for V.* Let us have another look at the earlier equations, replacing $\rho^2 - \rho^1$ by zero (due to the uniqueness of ρ just established):

$$\mathbb{E}_x v(X_t^1) - v(x) = \mathbb{E}_x \int_0^t h_1(X_s^1)\, ds.$$

Suppose $h_1 \ge 0$ and $\mu(x : \ h_1(x) > 0) > 0$. Due to the ergodicity, it follows that $\langle h_1, \mu_1 \rangle > 0$, which contradicts to the zero left hand side after division by t and $t \to \infty$. So, we conclude that

$$h_1 = 0, \quad \mu_1 - \text{a.s.}$$

Since $\mu_1 \sim \Lambda$ (see the Remark 1), we obtain that

$$\mathbb{E}_x v(X_t^1) - v(x) = 0,$$

due to Krylov's bound $\mathbb{E}_x \displaystyle\int_0^t h_1(X_s^1)\, ds \le N_t \|h_1\|_{L_2} = 0$.

Further, from

$$\mathbb{E}_x v(X_t^1) - v(x) = 0,$$

and due to the last statement of the Lemma 2 it follows that

$$v(x) = \lim_{t \to \infty} \mathbb{E}_x v(X_t^1) = \langle v, \mu_1 \rangle.$$

Hence, $v(x)$ **is a constant**. Recall that uniqueness of the first component V is stated up to a constant, and it was established that

$$v^1(x) - v^2(x) = \text{const}.$$

8. *Epsilon-approximation for ρ.* Denote $k(x) := F[\hat V, \hat\rho](x)$. Let (V, ρ) be the solution of the equation (6) and firstly let us apply Ito's formula with expectations to $(V - \hat V)(X_t^\alpha)$ where the strategy α is optimal for V:

$$\mathbb{E}V(X_t^\alpha) - \mathbb{E}_x \hat V(X_t^\alpha) - V(x) + \hat V(x) = \mathbb{E}_x \int_0^t (L^\alpha V - L^\alpha \hat V)(X_s^\alpha)\, ds$$

$$= \mathbb{E}_x \int_0^t (L^\alpha V + f^\alpha + \rho - L^\alpha \hat V - f^\alpha - \hat\rho)(X_s^\alpha)\, ds + (\hat\rho - \rho)t$$

$$= -\mathbb{E}_x \int_0^t (L^\alpha \hat V + f^\alpha + \hat\rho)(X_s^\alpha)\, ds + (\hat\rho - \rho)t$$

$$\le -\mathbb{E}_x \int_0^t \inf_u (L^u \hat V + f^u + \hat\rho)(X_s^\alpha)\, ds + (\hat\rho - \rho)t$$

$$= (\hat\rho - \rho)t - \mathbb{E}_x \int_0^t k(X_s^\alpha)\, ds.$$

Since the left hand side here is bounded for a fixed x, divide all terms of the latter inequality by t and let $t \to \infty$. Then due to the assumption $\inf k \ge -\epsilon$, we get,

$$\hat\rho - \rho \ge -\epsilon. \tag{21}$$

Secondly, let us apply Ito's formula with expectations to $(V - \hat V)(X_t^\alpha)$ where this time the strategy α is optimal for $\hat V$ (i.e. where the $\inf_u G^u[\hat V](x)$ is attained):

$$\mathbb{E}V(X_t^\alpha) - \mathbb{E}_x \hat V(X_t^\alpha) - V(x) + \hat V(x) = \mathbb{E}_x \int_0^t (L^\alpha V - L^\alpha \hat V)(X_s^\alpha)\, ds$$

$$= \mathbb{E}_x \int_0^t (L^\alpha V + f^\alpha + \rho - L^\alpha \hat V - f^\alpha - \hat\rho)(X_s^\alpha)\, ds + (\hat\rho - \rho)t$$

$$\ge \mathbb{E}_x \int_0^t [\inf_u (L^u V + f^u + \rho)(X_s^\alpha) - (L^\alpha \hat V + f^\alpha + \hat\rho)(X_s^\alpha)]\, ds + (\hat\rho - \rho)t$$

$$= (\hat\rho - \rho)t - \mathbb{E}_x \int_0^t k(X_s^\alpha)\, ds.$$

Since the term $\mathbb{E}V(X_t^\alpha) - \mathbb{E}_x \hat V(X_t^\alpha) - V(x) + \hat V(x)$ here is again bounded for a fixed x, divide once more all terms in the whole calculus by t and let $t \to \infty$. Then due to the assumption $\sup k \le \epsilon$, we get,

$$\hat{\rho} - \rho \le \epsilon. \tag{22}$$

So, by virtue of (21) and (22), the **last claim of the Theorem is proved.**

Remark 2. *It seems plausible that for the last statement of the Theorem we could have assumed that the function k – i.e., the function $F[\hat{V}, \hat{\rho}](x)$ in the last Theorem's claim – is small in some integral norms, too. Also, at the moment we do not know how the Theorem 1 could be possibly extended to a multi-dimensional case, although, it would be nice to have such an extension. Generally, there is a hope that results on average long-term control will find applications in the area related to long-term insurance and reinsurance mathematical models.*

References

[1] Ari Arapostathis. On the policy iteration algorithm for nondegenerate controlled diffusions under the ergodic criterion in book. In: *Optimization, control, and applications of stochastic systems*, Systems Control Found. Appl., 1–12. Birkhäuser/Springer, New York, 2012.

[2] Ari Arapostathis, Vivek S. Borkar, and Mrinal K. Ghosh. *Ergodic control of diffusion processes.* Encyclopedia of Mathematics and its Applications 143. Cambridge: Cambridge University Press, 2012.

[3] Vivek S. Borkar. *Optimal control of diffusion processes.* Harlow: Longman Scientific & Technical; New York: John Wiley & Sons, 1989.

[4] L.D. Brown and R. Purves. Measurable selections of extrema. *Ann. Stat.*, 1: 902–912, 1973.

[5] E.B. Dynkin and A.A. Yushkevich. *Upravlyaemye markovskie protsessy i ikh prilozheniya.* Moskva: "Nauka", 1975. (in Russian)

[6] I.V. Evstigneev. Measurable selection theorems and probabilistic control models in general topological spaces. *Math. USSR, Sb.*, 59(1): 25–37, 1988.

[7] Ronald A. Howard. *Dynamic programming and Markov processes.* New York-London: John Wiley &; Sons, Inc. and the Technology Press of the Massachusetts Institute of Technology, 1960.

[8] Ronald A. Howard. *Dynamic probabilistic systems. Vol. II: Semi-Markov and decision processes. Reprint of the 1971 original ed.* Mineola, NY: Dover Publications, 577-1108, 2007.

[9] Rafail Khasminskii. *Stochastic stability of differential equations. With contributions by G.N. Milstein and M.B. Nevelson. 2nd completely revised and enlarged ed.* Berlin: Springer, 2012.

[10] N.V. Krylov. *Lectures on elliptic and parabolic equations in Sobolev spaces.* Providence, RI: AMS, 2008.

[11] N.V. Krylov, On the selection of a Markov process from a system of processes and the construction of quasi-diffusion processes. *Math. USSR Izv.* 7: 691–709, 1973.

[12] A.A. Ljapunow, E.A. Stschegolkow, and W.J. Arsenin. *Arbeiten zur deskriptiven Mengenlehre.* Mathematische Forschungsberichte. 1. Berlin: VEB Deutscher Verlag der Wissenschaften, 1955.

[13] P. Mandl. *Analytical treatment of one-dimensional Markov processes.* Berlin-Heidelberg-New York: Springer Verlag, 1968.

[14] H. Mine and Shunji Osaki. *Markovian decision processes.* New York: American Elsevier Publishing Company, Inc., 1970.

[15] R. Morton. On the optimal control of stationary diffusion processes with inaccessible boundaries and no discounting. *J. Appl. Probab.*, 8: 551–560, 1971.

[16] E. Pardoux and A.Yu. Veretennikov. On the Poisson equation and diffusion approximation. I. *Ann. Probab.*, 29(3): 1061–1085, 2001.

[17] È. Pardoux and A.Yu. Veretennikov. On Poisson equation and diffusion approximation. II. *Ann. Probab.*, 31(3): 1166–1192, 2003.

[18] Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, 2005.

[19] E.A. Shchegol'kov. Über die Uniformisierung gewisser B-Mengen. *Dokl. Akad. Nauk SSSR, n. Ser.*, 59: 1065–1068, 1948.

[20] A.Yu. Veretennikov. On polynomial mixing for SDEs with a gradient-type drift. *Theory Probab. Appl.*, 45(1): 160–164, 2000.